

## THE STRUCTURE OF COLLAGEN

BY ALEXANDER RICH AND F. H. C. CRICK

Physical Chemistry Section, National Institute of Mental Health,  
Bethesda 14, Maryland, U.S.A.

MRC Unit for the Study of the Molecular Structure of Biological Systems,  
Cavendish Laboratory, Cambridge

*Received 4th February, 1957*

It is now generally agreed that there are only two structures, known as Collagen I and Collagen II, which can explain the wide-angle X-ray picture of collagen. The paper describes these two structures, which are closely related, in a simple manner. Collagen II is more satisfactory than Collagen I, but whether any appreciable amount of the latter is present in collagen remains to be discovered.

In this short article we shall describe the structures recently proposed for collagen. We shall consider only the small-scale structure, which gives the wide-angle X-ray pattern, and not the large-scale structure shown by the long spacings, both parallel and perpendicular to the fibre axis.

We shall omit the description of previous work on this problem and also the detailed experimental evidence on which the structures are based, as these points will be fully covered in a comprehensive paper which we shall be publishing elsewhere, and also to some extent by other papers in this volume. We should state that as far as X-ray work is concerned the present ideas on the structure of collagen have sprung from the work of Ramachandran and his colleagues (see references under Ramachandran) and that in addition to our own work contributions have been made by the group at Kings' College, London (see references under Cowan) and at the Massachusetts Institute of Technology (Cohen and Bear, 1953; Bear, 1955 and 1956).

It is now generally believed that it is possible to construct only two models which are compatible with the X-ray, infra-red, chemical and physico-chemical data (see other chapters in this volume for references). These we have called structure I and structure II (Rich and Crick, 1955). They have also been called "plus" and "minus" respectively by Ramachandran (1956), and "anti-clockwise" and "clockwise" respectively by Cowan, McGavin and North (1955).

All four groups of workers are now agreed that structure II is more satisfactory stereochemically and a better fit with the X-ray picture than structure I. However the X-ray picture is so poor that it is possible that stretches of both structure II and structure I exist, though the former probably predominates in stretched collagen.

Structures I and II are very similar. We shall first describe the features they have in common. Both consist of three separate polypeptide chains. Each chain is coiled into a helix, and these three helical chains slowly coil round one another. The easiest way to grasp the arrangement is to consider an imaginary derivation of the structure as follows. Take one polypeptide chain, ignoring for the moment the amino acid side-chains. Coil it so that it has a three-fold left-hand screw axis, such that the screw which takes one from one residue to the next has a rotation of  $-120^\circ$  and a displacement, in the fibre direction, of about 3 Å. This is the backbone proposed for poly-L-proline by Cowan and McGavin (1955) and for polyglycine II by Crick

and Rich (1955). The screw is chosen left-hand so that L-proline will be able to form part of the structure. Two such chains are shown side by side in figs. 1 and 2.

Put three such chains together side by side, with their axes parallel, to form a compact group of three chains, each being about 5 Å from the other

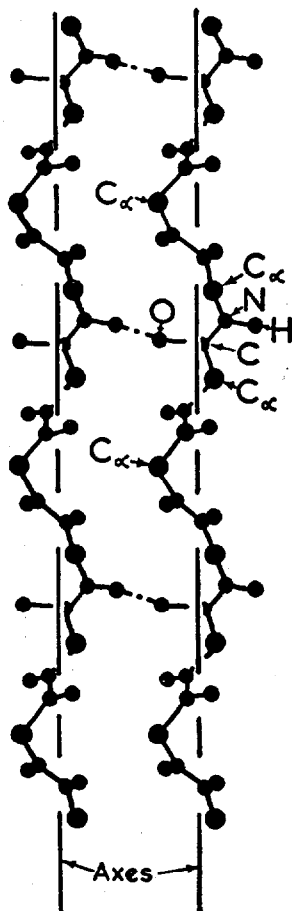


FIG. 1

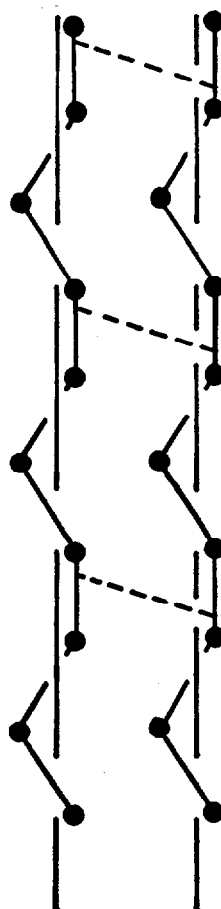


FIG. 2

FIG. 1.—Two polypeptide backbones shown side by side. It can be seen that each follows a helical path, having a left-hand 3-fold screw axis. (The axes are shown symbolically as vertical lines.) The broken lines between the two chains represent hydrogen bonds. The larger circles represent the  $C_\alpha$  carbon atoms.

FIG. 2.—A simplified version of fig. 1, in which only the  $C_\alpha$  carbon atoms are shown connected by short straight lines symbolizing the peptide groups. The broken lines represent hydrogen bonding.

two. The three chains should all be at the same level—that is, a suitable translation, *perpendicular* to the fibre axis, should take one from one chain to its neighbour (see figs. 3 and 4). It will be found that the three chains are also related by a three-fold screw axis running up the middle of the group of three.

Now consider one of the three chains of this group. Every third residue on this chain will be in an identical environment ; for example, every third

residue might be near the middle of the group of three chains, while the other two might be near the outside. If the chains are brought together in a suitable orientation it will be found that every third NH group, on the backbone of one chain, can make a hydrogen bond with every third CO

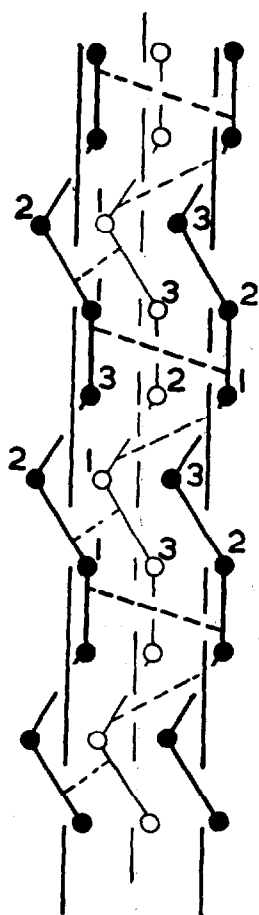


FIG. 3

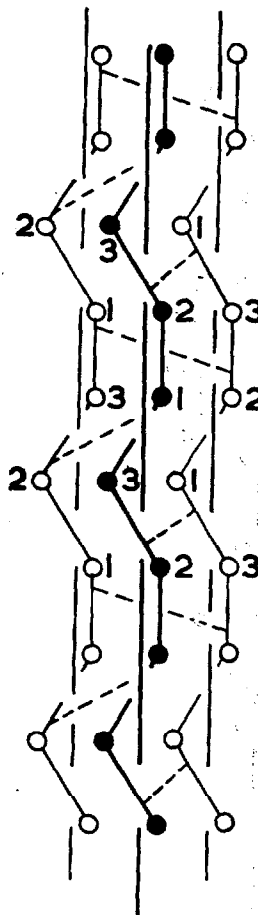


FIG. 4



FIG. 5

FIG. 3.—As fig. 2, but with a third polypeptide chain added *behind* the other two. This arrangement is related to Collagen I. The numbers 1, 2 and 3 represent the three types of side-chain positions. Broken lines represent hydrogen bonds.

FIG. 4.—As fig. 2, but with a third polypeptide chain added *in front* of the other two. This arrangement is related to Collagen II. The numbers 1, 2 and 3 represent the three types of side-chain positions. Broken lines represent hydrogen bonds.

FIG. 5.—Showing the general way in which the structures of figs. 3 and 4 are deformed to give the Collagen models. The solid lines represent the axes of the three polypeptide chains, which now follow gradual right-handed helices, instead of being straight and vertical. The broken line shows the common axis round which the three chains wind.

group on the backbone of a neighbouring chain. This is true for all three chains, since there is nothing in the arrangement which distinguishes between them. It is found by trial that there are only two orientations in which the three chains can be brought together to form satisfactory hydrogen bonds; one of these is related to structure I and the other to structure II. Thus the two structures differ mainly in the way the backbones of the three

chains are "phased" relative to one another. This can be seen by comparing figs. 3 and 4.

The actual models for collagen are derived from these two imaginary structures by deforming them so that the axes of the three chains (instead of running straight and parallel) twist slowly round one another in a gradual right hand helix as shown diagrammatically in fig. 5. The three-fold screw axis in the centre of the group is thus deformed so that its angle of rotation is  $-108^\circ$  (instead of  $-120^\circ$ ) and its translation in the fibre direction is  $2.86 \text{ \AA}$ , or up to about  $3 \text{ \AA}$  in stretched collagen.

The operation of this central screw axis takes one from a residue on one chain to a corresponding residue on the next chain. After *three* operations of this screw axis (giving therefore  $-324^\circ$  rotation and  $8.58 \text{ \AA}$  translation) one arrives back on the polypeptide chain from which one started, but three residues higher up. Since  $-324^\circ = +36^\circ$ , the screw axis relating every third residue on one chain is a right-hand rotation of  $+36^\circ$  and a translation of  $8.58 \text{ \AA}$ .

So far we have neglected the side-chains. If no side-chains were present it would be equally easy to build the two structures, but it is found from the study of scale models that the side-chains can be added much more satisfactorily to the backbone of structure II. As has been explained, every third residue finds itself in a similar environment as far as the polypeptide backbones are concerned. Thus there are three different types of position in which the side-chains can occur; we shall call these positions 1, 2 and 3. The restrictions on the placing of the side-chains for both structures are set out in the table, which should be studied carefully. It will be seen that the restrictions on structure I are rather severe, but these can be relaxed to some extent by quite small deformations, and for these reasons the allowed side-chains for a deformed structure I are also listed, though the table disguises the great complexity of the situations which arise once deformations are introduced.

TABLE 1—THE POSSIBLE POSITIONS OF SIDE-CHAINS

position	Collagen I		Collagen II
	undeformed	deformed	
1	glycine only	other residues may be possible: pro or hypro impossible	must be glycine
2	any residue including pro and hypro	any residue including pro and hypro	any residue including pro and hypro
3	glycine only	any residue, including pro and hypro, except valine	any residue, including pro and hypro
bonding of the OH of hypro in position 3	—	can make a hydrogen bond to the neighbouring chain within the group of three	sticks out radially away from the structure and cannot make a hydrogen bond <i>within</i> the group of three chains

The position of the hydrogen bonds between the three chains can be described as follows:

Collagen I: From the NH of the residue in position 1 (the "glycine" position) to the CO of the residue in position 1 on the neighbouring chain.

Collagen II: From the NH of the residue in position 1 (the "glycine" position) to the CO of the residue in position 2 (the "proline" position) on the neighbouring chain.

In structure I the N—H groups point anti-clockwise when viewed from the carboxyl ends of the chains ; in structure II clockwise.

It should be noted that the repeating sequence gly-pro-hydro, which the chemical evidence (Schroeder *et al.*, 1955 ; Kroner *et al.*, 1955) suggests is common in collagen, fits especially easily into Collagen II, though it is possible to fit it into Collagen I to some extent if some deformation of the basic structure is allowed.

The *position* of the hydro, however, is radically different in the two structures. In structure II it sticks out radially from the group of three chains, and it is possible that the OH group of the hydroxyproline could form a hydrogen bond with a neighbouring group of three chains. In Collagen I, on the other hand, it appears possible to make a hydrogen bond *within* the group of three chains, since the OH of the hydro lies at a suitable distance from the CO of one of the other two chains.

It thus becomes important to know whether the apparent stabilization of the collagen structure by hydroxyproline (Gustavson, 1953, 1955 and 1956 ; Takahashi and Tanaka, 1953) persists into dilute solutions in which the groups of three chains are too far apart from each other to be in contact.

The present position can therefore be summarized as follows : only two structures, closely related, appear possible for collagen. One of these, structure II, seems to fit the data for stretched collagen better than the other. What proportion, if any, of structure I exists in collagen, under different conditions, remains to be discovered.

- Bear, R. S. (1955). *Fibrous Proteins and Their Biological Significance*. Symp. No. IX. Soc. Exp. Biol. Cambridge University Press.
- Bear, R. S. (1956). *J. Biophysic. and Biochem. Cytol.* **2**, 363.
- Cohen, C. and Bear, R. S. (1953). *J. Amer. Chem. Soc.* **75**, 2783.
- Cowan, P. M. and McGavin, S. (1955). *Nature* **176**, 501.
- Cowan, P. M., McGavin, S. and North, A. C. T. (1955). *Nature* **176**, 1062.
- Cowan, P. M., North, A. C. T. and Randall, J. T. (1953). *The Nature and Structure of Collagen*, Butterworths, London.
- Cowan, P. M., North, A. C. T. and Randall, J. T. (1955). *Fibrous Proteins and Their Biological Significance*. Symp. No. IX. Soc. Exp. Biol. Cambridge University Press.
- Crick, F. H. C. and Rich, A. (1955). *Nature* **176**, 780.
- Gustavson, K. H. (1953). *Svensk. Kem. Tidskr.* **65**, 70.
- Gustavson, K. H. (1955). *Nature* **175**, 70.
- Gustavson, K. H. (1956). *The Chemistry and Reactivity of Collagen*. Academic Press, New York.
- Kroner, T. D., Tabroff, W. and McGarr, J. J. (1955). *J. Amer. Chem. Soc.* **77**, 3356.
- Ramachandran, G. N. (1956). *Nature* **177**, 710.
- Ramachandran, G. N. and Ambady, G. K. (1954). *Curr. Sci.* **23**, 349.
- Ramachandran, G. N. and Kartha, G. (1954). *Nature* **174**, 269.
- Ramachandran, G. N. and Kartha, G. (1955). *Nature* **176**, 593.
- Rich, A. and Crick, F. H. C. (1955). *Nature* **176**, 915.
- Schroeder, W. A., Kay, L. M., Le Gette, J., Honnen, L. and Green, R. C. (1954). *J. Amer. Chem. Soc.* **76**, 2783.
- Takahashi, T. and Tanaka, T. (1953). *Bull. Jap. Soc. Fish.* **19**, 603.